Lattices with non-Shannon Inequalities

Peter Harremoës
Copenhagen Business College
Copenhagen
Denmark
Email: harremoes@ieee.org

***Abstract*—We study the existence or absence of non-Shannon inequalities for variables that are related by functional dependencies. Although the power-set on four variables is the smallest Boolean lattice with non-Shannon inequalities there exist lattices with many more variables without non-Shannon inequalities. We search for conditions that ensures that no non-Shannon inequalities exist. It is demonstrated that 3-dimensional distributive lattices cannot have non-Shannon inequalities and planar modular lattices cannot have non-Shannon inequalities. The existence of non-Shannon inequalities is related to the question of whether a lattice is isomorphic to a lattice of subgroups of a group.**

***Index Terms*—Functional dependence, lattice, modularity, non-Shannon inequality, polymatroid.**

I. Introduction

The existence of non-Shannon inequalities have got a lot of attention since the first inequality was discovered by Z. Zhang and R. W. Yeung [1]. The basic observation is that any four random variables A, B, C, and D satisfy the following inequality

$$\begin{aligned}&2I\left(C;D\right)\leq\\&I\left(A;B\right)+I\left(A;C\uplus D\right)+3I\left(C;D\mid A\right)+I\left(C;D\mid B\right)\end{aligned}\tag{1}$$

where $I\left(A,B\right)$ denotes the mutual information between A and B and $I\left(C;D\mid B\right)$ denotes the conditional mutual information between C and D given B. Finally, $C\uplus D$ here denotes the random variable that takes values of the form (c,d) where $c=C$ and $d=D$. The inequality is non-Shannon in the sense that it cannot be deduced from inequalities of the form

$$\begin{aligned}H\left(X\uplus Y\right)&\geq&H\left(X\right)\\I\left(X;Y\mid Z\right)&\geq&0.\end{aligned}$$

Using that $I\left(X;Y\right)=H\left(X\right)+H\left(Y\right)$ and

$$\begin{aligned}&I\left(X;Y\mid Z\right)=\\&\quad H\left(X\uplus Z\right)+H\left(Y\uplus Z\right)-H\left(X\uplus Y\uplus Z\right)-H\left(Z\right)\end{aligned}$$

the last inequality can be rewritten as

$$H\left(X\uplus Z\right)+H\left(Y\uplus Z\right)\leq H\left(X\uplus Y\uplus Z\right)+H\left(Z\right),$$

which we will call the sub-modular inequality. Therefore the Shannon inequalities are the ones that can be deduced from using that entropy is non-negative, increasing and submodular. Later it was shown by F. Matus [2] that for four variables there exists infinitely many non-Shannon inequalities. It is easy to show that any inequality involving only three variables rather than four can be deduced from Shannon's inequalities. Now the power set of for variables is a Boolean algebra with 16 elements and any smaller Boolean algebra corresponds to smaller number of variables, so in a trivial sense the Boolean algebra with 16 elements is the smallest Boolean algebra for which there exists non-Shannon inequalities.

In the literature on non-Shannon inequalities all inequalities are expressed in terms of sets of variables and their joins. Another way to formulate this is that the inequalities are stated for the free $\uplus$-semilattice generated by a finite number of variables. In this paper we will also consider intersection of variables. We note that for sets of variables we have the inequality

$$I\left(X;Y\mid Z\right)\geq H\left(X\cap Y\mid Z\right).$$

This inequality have even inspired some authors to see the notation $I\left(\cdot\wedge\cdot\right)$ to denote mutual information.

Although non-Shannon inequalities have been known for more than a decade they have found remarkable few applications compared with Shannon's inequalities. One of the reasons for this is that there exists much larger lattices that a Boolean algebra with 16 elements. The simplest example is are the Markov chains.

$$X_1\to X_2\to X_3\to\cdots\to X_n$$

where X_1 determines X_2 which determines X_3 etc. For such a chain one has

$$H\left(X_1\right)\geq H\left(X_2\right)\geq H\left(X_3\right)\geq\cdots\geq H\left(X_n\right)\geq 0.$$

These inequalities are all instances of monotonicity of the entropy function, and it is quite clear that these inequalities are sufficient in the sense that for any sequence of values that satisfies these inequalities there exists random variables related by a deterministic Markov chain with these values as entropies.

In this paper we look at entropy inequalities for random variables that are related by functional dependencies. Functional dependencies gives an ordering of variables into a lattice. Such functional dependence lattices have

many applications in information theory, but in this short note we will focus on the question how one can detect whether a lattice of functionally related variables can non-Shannon inequalities. In particular we are interested in determination of the "smallest" lattice with non-Shannon inequalities. Here we should note that there are several ways of measuring the size of a lattice, and also note that in order to achieve interesting results have have to restrict our attention to special classes of lattices.

Non-Shannon inequalities have been studied using matroid theory but matroids are equivalent to atomistic semimodular lattices. For the study of non-Shannon inequalities it is more natural to look at general lattices rather than matroids because many important applications involve lattices that are not atomistic or not semimodular. For instance the deterministic Markov chain gives a lattice that is not atomistic. It is known that a function is entropic if and only if it can (approximately) equal to the logarithm of the index of a subgroup in a group. The lattice of subgroups of a Abelian group is modular and atomistic and can be described by matroid theory. A switch from matroids to lattices corresponds to a switch from Abelian groups to more general groups.

II. Lattices of functional dependence

Many problems in information theory and cryptography can be formulated in terms functional dependencies. For instance one might be interested in giving each member of a group part of a password in such a way that no single person can recover the whole password but any two members are able to recover the password. Here the password should be a function of the variables known by any two members but not a function of a variable hold by any single member. In this section we shall briefly describe functional dependencies and their relation to lattice theory. The relation between functional dependence and lattices has previously been studied [3], [4], [5], [6]. The relation between functional dependencies and Bayesian networks is described in [7].

Inspired by Armstrong's theory of relational databases [8] we say that a relation $\rightarrow$ in a lattice L satisfies *Armstrong's axioms* if it satisfies the following properties.

Transitivity If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$.

Reflexivity If $X \geq Y$, then $X \rightarrow Y$.

Augmentation If $X \rightarrow Y$, then $X \vee Z \rightarrow Y \vee Z$.

In a database $X \rightarrow Y$ should mean that there exists a function such that $Y = f(X)$ obviously satisfies these inference rules so as an axiomatic system it is sound. Armstrong proved that these axioms form a complete set of inference rules . That means that if a set A of functional dependencies is given and a certain functional dependence $x \rightarrow y$ holds in any database where all the functional dependencies in A hold then $x \rightarrow y$ holds in that database. Therefore for any functional dependence $x \rightarrow y$ that cannot be deduced using Armstrong's axioms the exist a database where the functional dependence is violated [9], [10]. As a consequence there exists a database where a functional dependence holds if and only if it can be deduced from Armstrong's axioms. A lattice element X is said to be closed if $X \rightarrow Y$ implies that $X \geq Y$. The smallest lattice element greater than X will be denoted $cl(X)$.

Theorem 1. *The set of closed elements form a lattice. For any finite lattice there exist a set of related variables such that the elements of the lattice corresponds to closed sets under functional dependence.*

According to the theorem any lattice can be considered as a closed set of variables under some functional dependence relation where $X \rightarrow Y$ if and only if $X \supseteq Y$. On the set of closed sets the meet operation is given by $X \cap Y$ and the join operation is given by $X \uplus Y = cl(X \cup Y)$. We observe that the set of closed sets is a subset of the original lattice that is closed under intersection. Such a subset is called a $\cap$-semilattice or a closure system. Any closure system defines a relation that satisfies Armstrong's axioms.

On a lattice *submodularity* of a function h is defined via the inequality $h(X) + h(Y) \geq h(X \uplus Y) + h(X \cap Y)$. A *polymatroid function* on a lattice can then be defined as a function that is non-negative, increasing and sub-modular. The relation $h(X \uplus Z) + h(Y \uplus Z) = h(X \uplus Y \uplus Z) + h(Z)$ defines a relation denoted $(X \perp Y \mid Z)$ that satisfies the properties:

Existence $(X \perp Y \mid X)$.

Symmetry $(X \perp Y \mid W)$ if and only if $(Y \perp X \mid W)$.

Decomposition If $(X \perp Y \uplus Z \mid W)$ then $(X \perp Z \mid W)$.

Contraction $(X \perp Z \mid W)$ and $(X \perp Y \mid Z \uplus W)$ implies $(X \perp Y \uplus Z \mid W)$.

Weak union $(X \perp Y \uplus Z \mid W)$ implies $(X \perp Y \mid Z \uplus W)$.

We say that a relation that satisfies these properties is *semi-graphoid*. Note that we allow the elements to have non-empty intersection. Note also that the existence property is normally not included in the list of semi-graphoid properties. If $(B \perp B \mid A)$ we write $A \supseteq_{\perp} B$. If h denotes the Shannon entropy H then $A \supseteq_{\perp} B$ simply means that $H(B \mid A) = 0$ or equivalently that B is almost surely a function of A.

Theorem 2. *If $(L, \cap, \uplus)$ is a lattice with a semi-graphoid relation $(\cdot \perp \cdot \mid \cdot)$ then the relation $\supseteq_{\perp}$ satisfies Armstrong's axioms. The relation $(\cdot \perp \cdot \mid \cdot)$ restricted to the lattice of closed lattice elements is semi-graphoid. If the semi-graphoid relation $(\cdot \perp \cdot \mid \cdot)$ is given by a polymatroid function h then h is also polymatroid if it is restricted the lattice of closed elements.*

III. Entropy in functional dependence lattices

Definition 3. A polymatroid function h on a lattice L is said to be entropic if there exists a function f form L into a set of random variables such that $h(x) = H(f(x))$ for any element in the lattice.

Let L denote a lattice and let $\Gamma(L)$ denote the set of polymatroid functions on L. Let $\Gamma^*(L)$ denote the set of entropic functions on L and let $\bar{\Gamma}^*(L)$ denote the closure of this set.

Definition 4. A lattice is said to be a *Shannon lattice* if any polymatroid function can be realized approximately by random variables, i.e. $\Gamma(L) = \bar{\Gamma}^*(L)$.

Both $\Gamma(L)$ and $\bar{\Gamma}^*(L)$ are polyhedral sets and often we may normalize the polymatroid functions by requiring that the value at the maximal element is 1. One may then check whether a lattice is a Shannon lattice by checking that the extreme polymatroid functions are entropic.

From the definition we immediately get the following result.

Proposition 5. *If L is a Shannon lattice and M is a subset that is a $\cap$-semilattice then M is a Shannon lattice. In particular all sub-lattices of a Shannon lattice are Shannon lattices.*

With these results at hand we can start hunting non-Shannon lattices. We take a lattice that may or may not be a Shannon lattice. We find the extreme polymatroid functions and for each extreme point we determine the lattice of closed elements using Theorem 2. Each of these lattices of closed sets have a much simpler structure than the original lattice and the goal is now to check if these lattices are Shannon lattices or not. It turns out that there are quite few of these reduced lattices and they could be considered as the building blocks for larger lattices.

The simplest lattice just has just two elements. The only normalized polymatroid function takes the values zero and one. It is obviously entropic.

We recall that an element i is $\uplus$-irreducible if $i = x \uplus y$ implies that $i = x$ or $i = y$. An $\cap$-irreducible element is defined similarly. An element is double irreducible if it is both $\uplus$-irreducible and $\cap$-irreducible. The lattice denoted M_n is a modular lattice with a smallest element, a largest element and $n-2$ double irreducible elements arranged in-between.

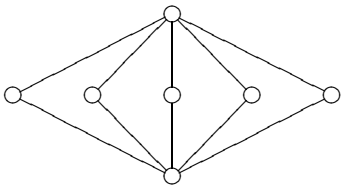

Figure 1. Hasse diagram of the lattice M_7.

Theorem 6. *For any n the lattice M_n is a Shannon lattice.*

Proof: The proof is essentially the same as the solution to the cryptographic problem stated in the beginning of Section II. The idea is that one should look for groups with a subgroup lattice M_n and then check that the subgroups of such group are actually have the right cardinality. ■

Corollary 7. *Any polymatroid function that only takes the values $0, 1/2$, and 1 is entropic.*

Proof: Assume that the polymatroid function h only takes the values $0, 1/2$, and 1. Then h defines a semi-graphoid relation and the closed elements form a lattice isomorphic to M_n for some integer n. The function h is entropic on M_n so h is also entropic on the original lattice. ■

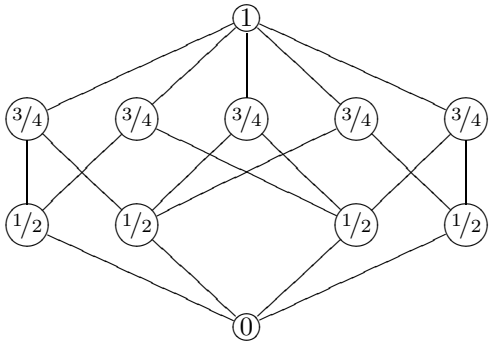

Figure 2. Lattice with a non-entropic polymatroid function.

The Boolean lattice with four atoms is the smallest non-Shannon Boolean algebra. Nevertheless there are smaller non-Shannon lattices. Figure 2 illustrates a lattice with just 11 elements that violates Inequality 1. This corresponds to the fact that the lattice in Figure 2 is not equivalent to a lattice of subgroups of a finite group. The lattices that are equivalent to lattices of subgroups of finite groups have been characterized [11], but the characterization is too complicated to describe in this short note. Using the ideas from [2] one can prove that the lattice in Figure 2 has infinitely many non-Shannon inequalities. Note that this lattice is atomistic but not semimodular, but it is lower locally distributive. Any semimodular lattice that contains the lattice in 2 as a $\cap$-semilattice also contains a power set on four points as a $\cap$-semilattice. The following lemma gives a considerable reduction in the number of inequalities that one has to consider in search for extreme polymatroid functions.

Lemma 8. *If h is submodular and increasing on $\cap$-irreducible elements then h is increasing.*

Theorem 9. *The lattice in figure 2 is the lower locally distributive non-Shannon lattice with fewest elements.*

Proof: There exists a nice presentation of lower locally distributive lattices [12] (In this paper the author works with the dual lattices). With this representation one it is relatively simple to create a list of all lower locally distributive lattices with 11 elements or fewer. Each lattice has finitely many extreme polymatroid functions. These can be found using the R program with package rcdd.

Each of these extreme polymatroid functions in each of these lattices has been checked to be entropic. ■

One may ask if there exists a lattice with fewer points than 11 that is non-Shannon.

Theorem 10. *Any lattice with 7 or fewer elements is a Shannon lattice.*

Proof: Up to isomorphism there only exist finitely many lattices with 7 or fewer elements. Each lattice has finitely many extreme polymatroid functions. These can be found using the R program with package rcdd. Each of these extreme polymatroid functions in each of these lattices has been checked to be entropic. ■

IV. Ingleton inequalities

The polymatroid function on Figure 2 does not only violate some non-Shannon inequalities, but it also violates an Ingleton inequality [13]. The Ingleton inequalities are inequities of the form

$$h(C)+h(D)+h(A \uplus C \uplus D)+h(B \uplus C \uplus D)+h(A \uplus B) \leq h(C \uplus D)+h(C \uplus A)+h(C \uplus B)+h(D \uplus A)+h(\uplus B).$$

A more instructive way of formulating the Ingleton inequalities is in terms of conditional mutual information.

$$I(X;Y \mid Z) \leq I(X;Y \mid Z \uplus V)+I(X;Y \mid Z \uplus W)+I(V;W \mid Z).$$

The Ingleton inequalities are satisfied for rank functions of representable matroid. In particular all entropic functions that can be described by Abelian groups satisfy the Ingleton inequalities. If a polymatroid on a lattice satisfies the Ingleton inequality the associated semi-graphoid relation satisfies the following property.

Strong contraction If $(X \perp Y \mid Z \uplus V)$ and $(X \perp Y \mid Z \uplus V)$ and $(V \perp W \mid Z)$ then $(X \perp Y \mid Z)$.

Like the Ingleton inequality strong contraction does not hold for all entropic polymatroid functions, but it does hold for most graphical models of independence like Bayesian networks. Recently it was demonstrated that strong contraction is essential for giving a lattice characterization of an certain system of inference rules for conditional independence [14]. In [14] strong contraction was used in conjunction with the following property.

Strong union If $(X \perp Y \mid Z)$ then $(X \perp Y \mid Z \uplus W)$.

Strong union is a quite restrictive condition, but it does hold for Markov chains and other Markov networks. The entropy inequality corresponding to strong union is

$$I(X;Y \mid Z) \leq I(X;Y \mid Z \uplus W).$$

If a polymatroid function satisfies the strong union inequality we get a significant reduction in the complexity of the problem.

Computer experiments support the following conjecture.

Conjecture 11. *If a polymatroid function on a lattice satisfies the Ingleton inequalities and the strong union inequalities then the function is entropic.*

It is worth noting that in [14] the authors use a lattice technique that is slightly different from the one developed in the present paper.

V. Distributive and modular lattices

The power-set of four variables is a distributive lattice so one may ask if there exists any distributive lattice with non-Shannon inequalities without this Boolean lattice as sub-lattice. We recall that a lattice is said to be *modular* if $a \subseteq b$ implies that

$$a \uplus (x \cap b) = (a \uplus x) \cap b$$

for any lattice element x. For modular lattices the following lemma gives a considerable reduction in the number of inequalities that one has to consider in the search for extreme points.

Lemma 12. *Let L be a modular lattice with a function h that is submodular on any sub-lattice with elements $a, b, a \cap b$ and $a \uplus b$ where $a \cap b$ is covered by a and b. Then the function h is submodular on L.*

For a distributive lattice the order dimension equals the maximal number of $\uplus$-irreducible elements (or maximal number of $\cap$-irreducible elements) needed in a decomposition of an element in the lattice. Distributive lattices may also be represented as ideals in partially ordered sets and the order dimension is also equal to the maximal anti-chain in the partially ordered set used in such a representation.

Theorem 13 ([15])**.** *Let L be a distributive lattice. Then L can be embedded as a sub-lattice into the n-th power of a chain if and only if it has order dimension at most n.*

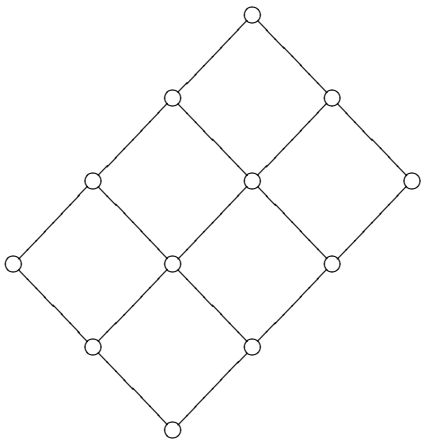

Figure 3. A product of two chains.

Theorem 14. *A distributive lattice is Shannon if and only if the order dimension at most 3.*

The free distributive lattice with three generators is a lattice on with the property that any distributive lattice generated by three elements is isomorphic to a sub-lattice. The free distributive lattice with three generators has 18 elements [16, 45-46, Theorem 10] and is three dimensional. Therefore we get the following result.

Corollary 15. *Any distributive lattice with 3 generators is a Shannon lattice.*

With three generators one can also define the free modular lattice. This lattice has 28 elements [16, 46-47, Theorem 11] and by explicit calculations one can check that it is a Shannon lattice.

Proposition 16. *The free modular lattice with 3 generators is a Shannon lattice.*

If we do not require that the lattice is modular (or belong to some other nice lattice variety) the result does not hold. The free lattice with three elements contain a sub-lattice isomorphic with the four dimensional Boolean algebra that is not a Shannon lattice. Therefore it would be interesting to know if there exists larger lattice varieties that the variety of modular lattices for which a free lattice with three generators in the variety is a Shannon lattice.

Theorem 17. *Any modular planar lattice is a Shannon lattice.*

The proof uses that it has it was recently proved that a planar modular lattices can be represented as a distributive lattice with a number of double irreducible elements added [17] as illustrated in Figure 4. Each of the extreme polymatroid functions on a planar modular lattice corresponds to a complicated cryptographic protocol or secrecy sharing scheme.

Acknowledgment

I want to thank Søren Riis and Sune Jacobsen for useful discussion during a research stay at Queen Mary University in January 2012.

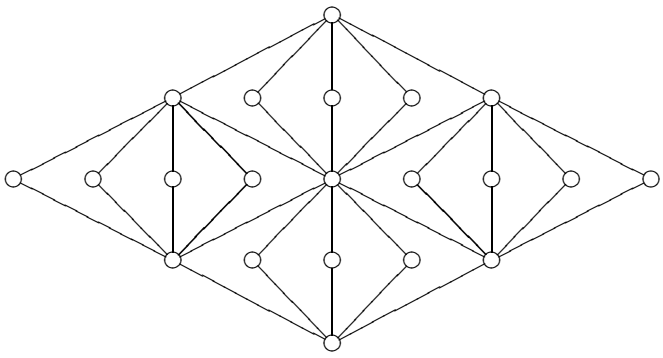

Figure 4. A planar modular lattice.

References

[1] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans . Inform. Theory*, vol. 44, no. 4, pp. 1440–1452, July 1988.

[2] F. Matús, "Infinitely many information inequalities," in *Proc. International Symposium on Information Theory (ISIT) 2007. Nice France*, June 2007, pp. 2101–2105.

[3] J. Demetrovics, L. Libkin, and I. B. Muchnik, "Functional dependencies and the semilattice of closed classes," in *MFDBS '89 Proceedings of the 2nd Symposium on Mathematical Fundamentals of Database Systems*. Springer, 1989, pp. 136–147.

[4] ——, "Functional dependencies in relational databases: A lattice point of view," *Discrete Applied Mathematics*, vol. 40, no. 2, pp. 155–185, Dec. 1992.

[5] M. Levene, "A lattice view of functional dependencies in incomplete relations," *Acta Cybernetica*, vol. 12, pp. 181–207, 1995.

[6] P. Harremoës, "Functional dependences and Bayesian networks," in *Proceedings WITMSE 2011*, Helsinki, 2011. [Online]. Available: http://www.harremoes.dk/Peter/FunctionalWITSME.pdf

[7] ——, "Influence diagrams as convex geometries," 2015, submitted. [Online]. Available: www.harremoes.dk/Peter/FunctionalDAG.pdf

[8] W. W. Armstrong, "Dependency structures of data base relationships," in *IFIP Congress*, 1974, pp. 580–583.

[9] J. D. Ullman, *Principles of Database and Knowledge-base Systems*. Stanford: Computer Science Press, 1989, vol. 1.

[10] M. Levene and G. Loizou, *A Guide Tour of Relational Databases and Beyond*. Springer, 1999.

[11] R. Schmidt, *Subgroup Lattices of Groups*. Walter de Gruyter, 1994.

[12] G. Behrendt, "Representations of locally distributive lattice," *Portugaliae Mathematica*, vol. 48, no. 3, pp. 351–355, 1991.

[13] L. Guille, T. Chan, and A. Grant, "The minimal set of Ingleton inequalities," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1849–1864, April 2011.

[14] M. Niepert, M. Gyssens, B. Sayrafi, and D. V. Gucht, "On the conditional independence implictation problem: A lattice-theoretic approach," *Artificial Intelligence*, vol. 202, pp. 29–51, 2013.

[15] R. P. Dilworth, "A decomposition theorem for partially ordered sets," *Ann. of Math.*, vol. 51, no. 2, pp. 161–166, 1950.

[16] G. Grätzer, *Lattice Theory*. Dover, 1971.

[17] G. G. W. Quackenbush, "The variety generated by planar modular lattices," *Algebra universalis*, vol. 63, no. 2-3, pp. 187–201, 2010.

[18] G. Grätzer, *General Lattice Theory*, second edition ed. Birkhäuser, 2003.

[19] N. Caspard and B. Monjardet, "The lattices of closure systems, closure operators, and implicational systems on a finite set: a survey," *Discrete Applied Mathematics*, vol. 127, no. 2, pp. 241–269, 2003.

[20] G. Paolini, "Independence logic and abstract independence relations," Sept. 2014, to appear in Mathematical Logic Quarterly. [Online]. Available: http://arxiv.org/pdf/1401.6907.pdf

[21] T. H. Chan and R. W. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1992–1995, 2002.

This appendix contains two sections with a more careful description of the relation between functional dependencies, lattices, semi-graphoid relations and polymatroid functions. The appendix also contains proofs of some of the theorems stated in the paper. In order to keep within this note short some proofs have been foreshortened or have been omitted.

Appendix A
The functional dependence lattices

In this section we shall describe functional dependencies and relate it to lattice theory. Much of the terminology is taken from database theory. The relation between functional dependence and lattices has previously been studied [3], [4], [5].

First we shall consider a set of attributes/variables V_i and subsets of this set of variables. Each attribute V_i takes values in some set W_i. The set of subsets is also called the power set and is ordered by inclusion. With this ordering the power set is a lattice with intersection and union as lattice operations. We note that the smallest element in the lattice is the empty set $\emptyset$ and the largest element is the whole set. One consider a set of tuples (records) that all share the same attributes. A *relation* R is a set of tuples and an assignment of a value in W_i to each attribute V_i. One may think of a relation as a function from tuples to the product space $\prod V_i$. If X and Y are sets of attributes we say that Y functionally dependence on X in the relation R and write $X \to Y$ if $\prod_{i\in X} V_i(t_1) = \prod_{i\in X} V_i(t_2)$ implies $\prod_{i\in Y} V_i(t_1) = \prod_{i\in Y} V_i(t_2)$.

Inspired by Armstrong's theory of relational databases we say that a relation $\to$ in a lattice L satisfies *Armstrong's axioms* if it satisfies the following properties.

Transitivity If $X \to Y$ and $Y \to Z$, then $X \to Z$.

Reflexivity If $X \geq Y$, then $X \to Y$.

Augmentation If $X \to Y$, then $X \vee Z \to Y \vee Z$.

Functional dependence $\to$ in a database obviously satisfies these inference rules so as an axiomatic system it is sound. Armstrong proved that these axioms form a complete set of inference rules. That means that if a set A of functional dependencies is given and a certain functional dependence $x \to y$ holds in any database where all the functional dependencies in A hold then $x \to y$ holds in that database. Therefore for any functional dependence $x \to y$ that cannot be deduced using Armstrong's axioms the exist a database where the functional dependence is violated [9], [10]. As a consequence there exists a database where a functional dependence holds if and only if it can be deduced from Armstrong's axioms.

Theorem 18. *A relation $\to$ on the elements of a lattice satisfies Armstrong's axioms if and only if $\to$ is a preordering that satisfies the following two properties.*

Decomposition *If $Z \to X \vee Y$, then $Z \to X$ and $Z \to Y$.*

Union *If $Z \to X$ and $Z \to Y$, then $Z \to X \vee Y$.*

Proof: Assume that $\to$ satisfies Armstrong's axioms. Then $X \geq X$ implies $X \to X$ so that $\to$ is reflexive. To prove the union property assume that $Z \to X$ and $Z \to Y$. Then $Z \vee Z \to X \vee Z$ and $X \vee Z \to X \vee Y$ by augmentation. Then transitivity implies $Z \to X \vee Y$. To prove the decomposition property assume that $Z \to X \vee Y$. In the lattice we have $X \vee Y \geq X$ and by reflexivity $X \vee Y \to X$. Now transitivity implies $Z \to X$. In the same way it is proved that $Z \to Y$.

Assume that $\to$ is a preordering that satisfies decomposition and union. To prove reflexivity assume that $X \geq Y$. Then $X \to X \vee Y$, which according to the decomposition property implies $X \to Y$. To prove augmentation assume that $X \to Y$. We have $X \vee Z \to X$ which together with transitivity implies $X \vee Z \to Y$. By reflexivity we have $X \vee Z \to Z$.Therefore the union property implies that then $X \vee Z \to Y \vee Z$. ∎

The first half of this proof was essentially given by Armstrong.

Let L denote a lattice with a relation $\to$ such that Armstrong's axioms are satisfied. For simplicity assume that L is finite. For $X \in L$ define $cl(X)$ as $\bigvee Y_i$ where the join is taken over all Y_i such that $X \to Y_i$. The union property implies that $cl(X)$ is maximal in the set of variables determined by X. With these definitions we see that $X \to Y$ if and only if $cl(X) \geq cl(Y)$. For a relation that satisfies Armstrong's axioms the unary operator cl satisfies the following conditions:

Extensivity $X \leq cl(X)$.

Isotony $X \leq Y$ implies $cl(X) \leq cl(Y)$.

Idempotens $cl(X) = cl(cl(X))$.

An unary operator that satisfies these three properties is called a *closure operator*. We say that X is closed if $cl(X) = X$. If X and Y are closed for some closure operator cl then it is easy to prove that $X \wedge Y$ is closed [18, Lemma 28]. A subset of a lattice that is closed under the meet operation, is called a semi-lattice or a *closure system.* In [19] closure systems were studied in more detail in the case where the lattice is a power set. The elements of the closure system are closed elements under the closure operator defined by $cl(\ell) = \bigwedge_{x\geq \ell, x\in A} x$.

Proposition 19. *Let $(L, \leq)$ denote a finite lattice. Assume that a subset A of L is closed under the meet operation. Then A is a lattice under the ordering $\leq$.*

Proof: The set A is partially ordered by $\leq$ so we just have to prove that any pair of elements in A has a least upper bound and a greatest lower bound. The greatest lower bound of $x, y \in A$ is $x \wedge y$. The least upper bound of x and y is $\bigwedge_{x\vee y\leq z, z\in A} z$. ∎

The lattice operations in A are given by $X \wedge_A Y = X \wedge Y$ and $X \vee_A Y = cl(X \vee Y)$. In particular the closed elements of a functional dependence relation form a lattice, and this was essentially the main result of Armstrong although he did not use lattice terminology. The converse of Armstrong's results is also true:

Theorem 20. *Let $(L, \leq)$ denote a finite lattice with a closure system A. Then the relation $x \to y$ is defined by $cl(x) \geq cl(y)$ satisfies Armstrong's axioms.*

Proof: It is easy to see that $\to$ defines a preordering. The union property is proved as follows. Assume that $x \to y$ and $x \to z$. Then cl(x) $\geq$cl(y) $\geq$y and $cl(x) \geq cl(z) \geq z$.Hence $cl(x) \geq y \vee z$ and $cl(x) \geq cl(y \vee z)$ so that $x \to y \vee z$. The decomposition property is proved by reversing the argument. ■

The theorem as it is formulated here probably appear somewhere in the literature on lattices although the author has not been able to locate a good reference.

Example 21. We consider three variables a, b and c that denote real numbers. Assume that $c = (a+b)^2$. Then the associated lattice is the lattice that is normally called S_7. This is illustrated in Figure 5.

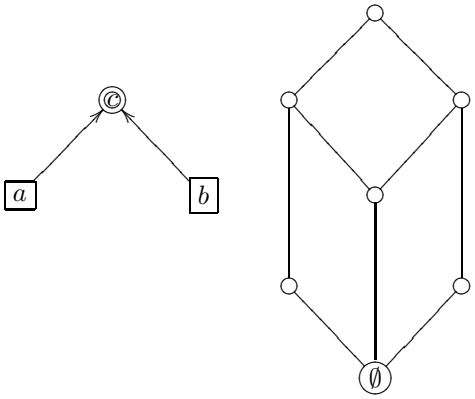

Figure 5. To the left and influence diagram for three variables is drawn with arrows indicating the direction of influence. To the right the Hasse diagram for the corresponding lattice of functional dependence is drawn with the smallest element ($\emptyset$) indicated. The name of this lattice is S_7.

Even simple examples of functional dependence lattices may be complicated to describe if they are not based on simple causal relations between the variables.

Example 22. This example concern fruit from a supermarket. Variable X tells whether the supermarket will sell it at normal price, or at a reduced price because it is close to the expiration date, or whether it is through out because the expiration date has been exceeded. Variable Z describes whether the fruit tastes very fresh, is eatable, or looks disgusting. The variable Y tells whether the fruit will make you sick or not. The functional dependencies are given by $Z \subseteq Y$ and $X \uplus Y = X \uplus Z$. The lattice is N_5. This is the standard example of a lattice that is not modular.

Theorem 23. *Any finite lattice can be represented as a closure system on a power set*

Proof: Let L be a lattice. For each $a \in L$ the principle ideal of a is $\downarrow (a) = \{x \in L \mid x \leq a\}$. This gives an

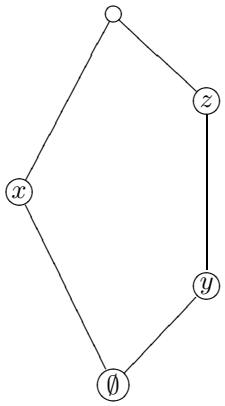

Figure 6. Hasse diagram of the lattice N_5.

embedding of L into the power set of L in such a way that meet in the lattice corresponds to intersection in the power set. ■

As a result any lattice is equivalent to a lattice of functional dependence, so all what can be said about functional dependence can be expressed in the language of lattices. Most of the time we will formulate our results in terms of closure systems. Since the notation for inclusion and intersection is fixed, we will use $\supseteq$ to denote the ordering of a functional dependence lattice and $\cap$ to denote the meet operation. If the lattice is the whole power set, i.e. a Boolean lattice then we will use $\cup$ to denote the join operation. If we have not assumed that the lattice is Boolean we may use $\vee$ or $\uplus$ or $\sqcup$ or some similar symbol to denote the join operation.

With the above results we can prove that Armstrong's axioms form a complete set of inference rules for functional dependencies.

Theorem 24. *For any finite lattice there exist a set of related variables such that the elements of the lattice corresponds to closed sets under functional dependence.*

Proof: A lattice can be represented as a closure system of a power set of some set I. To each element $i \in I$ we associate a binary variable V_i with values in $W_i = \{0, 1\}$. Let C denote the closed sets in the power sets. For each $c \in C$ we assign an tuple t_c so that

$$V_i(t_c) = 1 \text{ if } i \in c.$$

Assume that $a \supseteq b$. If $(V_i(t_{c_1}))_{i \in a} = (V_i(t_{c_2}))_{i \in a}$ for all tuples t_{c_j} then $(V_i(t_{c_1}))_{i \in b} = (V_i(t_{c_2}))_{i \in b}$ for holds for all tuples t_{c_j}. Hence $a \to b$.

Assume that $a \to b$. According to the definition it means that if $(V_i(t_{c_1}))_{i \in a} = (V_i(t_{c_2}))_{i \in a}$ for all tuples t_{c_j} then $(V_i(t_{c_1}))_{i \in b} = (V_i(t_{c_2}))_{i \in b}$ for holds for all tuples t_{c_j}. Assume that $(V_i(t_{c_1}))_{i \in a} = (V_i(t_{c_2}))_{i \in a}$. Then for all $i \in a$ we have $V_i(t_{c_1}) = V_i(t_{c_2})$ which is equivalent to $a \cap c_1 = a \cap c_2$. Similarly $(V_i(t_{c_1}))_{i \in b} = (V_i(t_{c_2}))_{i \in b}$ is equivalent to $b \cap c_1 = b \cap c_2$. Choose $c_1 = a$ and

$c_2 = a \uplus b$. Then $a \cap c_1 = a \cap c_2$ is automatically fulfilled and $b \cap c_1 = b \cap c_2$ can be rewritten as $b \cap a = b \cap (a \uplus b) = b$, which implies that $a \supseteq b$. ∎

Appendix B
Independence in lattices

In statistics one studies the relation $(A \perp B \mid C)$ meaning that A and B are independent given $\check{s}C$, where A, B and C are disjoint subsets of a set M of random variables with respect to a probability measure. We will call this notion of independence *statistical independence*.

We shall say that a relation $(\cdot \perp \cdot \mid \cdot)$ on a lattice $(L, \cap, \uplus)$ is a *semi-graphoid relation*, if it satisfies the following axioms:

Existence $(X \perp Y \mid X)$.

Symmetry $(X \perp Y \mid W)$ if and only if $(Y \perp X \mid W)$.

Decomposition If $(X \perp Y \uplus Z \mid W)$ then $(X \perp Z \mid W)$.

Contraction $(X \perp Z \mid W)$ and $(X \perp Y \mid Z \uplus W)$ implies $(X \perp Y \uplus Z \mid W)$.

Weak union $(X \perp Y \uplus Z \mid W)$ implies $(X \perp Y \mid Z \uplus W)$.

These properties should hold for all $X, Y, Z, W \in L$. We note that statistical independence with respect to a probability measure is semi-graphoid. In this paper we are particularly interested in the case where the subsets are not disjoint. A relation that satisfies the last for properties for disjoint sets in a power was said to be semi-graphoid [**?**]. In a recent paper [20] a much longer list of axioms for the notion of independence was given. Most of those axioms can be proved from the axioms stated in this paper.

Theorem 25. *A semi-graphoid relation $(\cdot \perp \cdot \mid \cdot)$ satisfies the following properties.*

Reflexivity *For all A we have $(X \perp X \mid X)$.*

Normality *If $(X \perp Y \mid W)$ then $(X \perp Y \uplus W \mid W)$.*

Monotonicity *If $(X \perp Y \mid W)$ and $Y \uplus W \supseteq Z$ then $(X \perp Z \mid W)$.*

Triviality $(X \perp \emptyset \mid Y)$

Base monotonicity *If $(A \perp B \mid D)$ and $B \supseteq C \supseteq D$ then $(A \perp B \mid C)$.*

Transitivity *If $(A \perp B \mid C)$ and $(A \perp C \mid D)$ and $B \supseteq C \supseteq D$ then $(A \perp B \mid D)$.*

Autonomy *If $(A \perp A \mid C)$ then $(A \perp B \mid C)$.*

In a power set of random variables we note that if A is independent of A given C then A is a function of C almost surely. If $(B \perp B \mid A)$ we write $A \supseteq_\perp B$.

Definition 26. An semi-graphoid relation is said to be consistent with $\subseteq$ if $X \subseteq Y$ is equivalent to $(X \perp X \mid Y)$.

Theorem 27. *If $(L, \cap, \uplus)$ is a lattice with a semi-graphoid relation $(\cdot \perp \cdot \mid \cdot)$ then the relation $\supseteq_\perp$ satisfies Armstrong's axioms. The relation $(\cdot \perp \cdot \mid \cdot)$ restricted to the lattice of closed lattice elements is semi-graphoid.*

Proof: **Reflexivity of** $\supseteq_\perp$ This follows according to the reflexivity property of $\perp$.

Transitivity Assume that $X \supseteq_\perp Y$ and $Y \supseteq_\perp Z$. Autonomy implies that $(Z \perp Z \uplus X \mid Y)$ and by weak union $(Z \perp Z \mid Y \uplus X)$. Autonomy and $X \supseteq_\perp Y$ together imply that $(Y \perp Z \mid X)$. Contraction then implies $(Z \perp Y \uplus Z \mid X)$. Decomposition gives $(Z \perp Z \mid X)$.

Decomposition This follows from the decomposition property of $\perp$.

Union Assume that $X \supseteq_\perp Y$ and $X \supseteq_\perp Z$. Then $(Y \perp Y \mid X)$ and $(Z \perp Z \mid X)$ and by autonomy $(Y \perp Y \uplus Z \mid X)$ and $(Z \perp Y \uplus Z \uplus Y \mid X)$. Hence $(Z \perp Y \uplus Z \mid Y \uplus X)$ by weak union and $(Y \uplus Z \perp Y \uplus Z \mid X)$ by contraction.

For the last result one just has to prove that $(X \perp Y \mid Z)$ if and only if $(X \perp cl_\perp(Y) \mid Z)$ if and only if $(X \perp Y \mid cl_\perp(Z))$. This follows from Armstrong's results. ∎

The significance of this theorem is that if we start with a semi-graphoid relation on a lattice then this semi-graphoid relation is also semi-graphoid when restricted elements that are closed under functional dependence.

Theorem 28. *Any finite lattice can be represented as a closure system of an semi-graphoid relation defined on a power-set.*

Proof: For any finite lattice L one identify the elements by sets of binary variables v_i, and a relation can be defined where the tuples have the form $i_c, c \in L$ as in the proof of Theorem 24. Each tuple can be identified with a point in the product space $\prod W_i$. Assign a uniform distribution to each point in the product space. With respect to this probability measure $(b \perp b \mid a)$ if and only if a determines b almost surely. Since the probability measure is discrete $(b \perp b \mid a)$ is valid if and only if $a \supseteq b$. ∎

The semi-graphoid relation defined in the proof of the previous theorem is based on the uniform distribution on the tuples. We note that any other distribution that has positive probability on the same tuples will also give a representation of the lattice in terms of a semi-graphoid relation. For disjoint sets independence will depend on the choice of probability distribution.

Appendix C
Proof of Proposition 5

Assume that L is a Shannon lattice and that M is a sublattice. Let $h : M \to \mathbb{R}$ denote a polymatroid function. For $\ell \in L$ let $\tilde{\ell}$ denote the $m \in M$ that minimize $h(m)$ under the constraint that $m \supseteq \ell$. Define the function $\tilde{h}(\ell) = h\left(\tilde{\ell}\right)$. Now $\tilde{h}$ is an extension of h and with this definition $\tilde{h}$ is non-negative and increasing. For $x, y \in L$ we have

$$\begin{aligned} \tilde{h}(x) + \tilde{h}(y) &= h(\tilde{x}) + h(\tilde{y}) \\ &\geq h(\tilde{x} \uplus \tilde{y}) + h(\tilde{x} \cap \tilde{y}) \\ &\geq \tilde{h}(x \uplus y) + \tilde{h}(x \cap y) \end{aligned}$$

because $\tilde{x} \uplus \tilde{y} \geq x \uplus y$ and $\tilde{x} \cap \tilde{y} \geq x \cap y$. Hence $\tilde{h}$ is submodular. By the assumption $\tilde{h}$ is entropic so the restriction of $\tilde{h}$ to M is also entropic.

Appendix D
Proof of Lemma 8

Assume that h is submodular and increasing on $\cap$-irreducible elements. We have to prove that if $a \supseteq c$ then $h(a) \geq h(c)$. In order to obtain a contradiction assume that c is a maximal element such that there exist an element a such $a \supseteq c$ but $h(a) < h(c)$. We may assume that a cover c. Since h is increasing at $\cap$-irreducible elements c cannot be $\cap$-ireducible. Therefore there exists a maximal element b such that $b \supseteq c$ but $b \not\supseteq a$. Since a cover c we have $a \cap b = c$. According to the assumptions $h(a) + h(b) \geq h(a \uplus b) + h(a \cap b)$ and $h(a \uplus b) \geq h(b)$ because c is a maximal element that violates monotonicity. Therefore $h(a) \geq h(a \cap b) = h(c)$.

Appendix E
Proof of Theorem 6

Let the values in the double irreducible elements be denoted $x_1, x_2, \ldots, x_{n-2}$. If $n = 1$ the extreme polymatroid functions are $x_1 = 0$ and $x_1 = 1$ and these points are obviously entropic. If $n = 4$ the extreme points are $(x_1, x_2) = (0, 1)$ and $(x_1, x_x) = (1, 0)$ and $(x_1, x_2) = (1, 1)$,which are all entropic.

Assume $n \geq 5$. Then the values should satisfy the inequalities

$$\begin{aligned} 0 \leq \quad x_i \quad &\leq 1 \\ x_i + x_j \quad &\geq \quad 1. \end{aligned}$$

If $(x_1, x_2, \ldots, x_{n-2})$ is an extreme point then each variables should satisfy one of the inequalities with equality. Assume $x_i = 0$. Then sub-modularity implies that $x_j = 1$ for $j \neq i$. The extreme point $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ is obviously entropic. If $x_i = 1$ this gives no further constraint on the other values, so it corresponds to an extreme point on a lattice with one less variable. Finally assume that $x_i + x_j = 1$ for all i, j. Then $x_i = 1/2$ for all i.

We have to find $n-2$ random variables $X_1, X_2 \ldots, X_{n-2}$ that are independent but such that any two determine the rest. Let p denote a prime larger than $n - 2$. Let Y and Z denote independent random variables with values in $\mathbb{Z}_p$ each with a uniform distribution. If X_j is defined to be equal to $Y + jZ$ then the variables X_j are mutually independent and any pair of these random variables determine all the other variables.

Instead of constructing the variables $X_1, X_2, \ldots, X_{n-2}$ we can find a group G and subgroups $G_1, G_2, \ldots, G_{n-2}$such that $|G| = 2|G_i|$ using general results about entropy inequalities and groups [21]. The group G can be chosen as $\mathbb{Z}_p \times \mathbb{Z}_p$where p is some prime number greater than $n - 2$. The group G has $p + 1$ subgroups isomorphic to $\mathbb{Z}_p$.

Appendix F
Proof of Lemma 12

Let a and b denote two lattice elements. We have to prove that $h(a) + h(b) \geq h(a \uplus b) + h(a \cap b)$.

Assume that $x_1, x_2, \ldots, x_n$ is sequence of elements such $a \cap b \subseteq x \subseteq x_2 \subseteq \cdots \subseteq x_n \subseteq a$. Define $y_i = x_i \uplus b$. Then modularity implies

$$\begin{aligned} x_{i+1} \cap y_i &= x_{i+1} \cap (b \uplus x_i) \\ &= (x_i \cap b) \uplus (x_{i+1} \cap x_i) \\ &= (a \cap b) \uplus x_i \\ &= x_i. \end{aligned}$$

We also have

$$\begin{aligned} x_{i+1} \uplus y_i &= x_{i+1} \uplus (b \uplus x_i) \\ &= (x_{i+1} \uplus x_i) \uplus b \\ &= x_{i+1} \uplus b \\ &= y_{i+1}. \end{aligned}$$

Assume that the modular inequality holds for all the sublattices $L_i = \{x_i, x_{i+1}, y_i, y_{i+1}\}$. Then we can add all the inequalities $h(x_{i+1}) + h(y_i) \leq h(x_i) + h(y_{i+1})$ to get $h(x_n) + h(y_1) \leq h(x_1) + h(y_n)$. Note that we can choose the sequence $x_1, x_2, \ldots, x_n$ such that x_{i+1} covers x_i and such that $x_1 = a \cap b$ and $x_n = a$. Therefore we it is sufficient to prove that $h(a) + h(b) \geq h(a \uplus b) + h(a \cap b)$ when a cover $a \cap b$.

Similarly it is sufficient to prove that $h(a) + h(b) \geq h(a \uplus b) + h(a \cap b)$ when b cover $a \cap b$. If a and b both cover $a \cap b$ then $M = \{a, b, a \cap b, a \uplus b\}$ is a sub-lattice of x^+ if $x = a \cap b$.

Appendix G
Proof of Theorem 14

If the lattice is one-dimensional we just get a deterministic Markov chain for which positivity and monotonicity are sufficient conditions for a function to be entropic. Assume that the lattice is two-dimensional.

We will show that an extreme polymatroid function only takes the values 0 and 1. Assume that $(A, \leq)$ the poset of irreducible elements in the distributive lattice. The proof is by induction on the number of elements k in the lattice. If $k = 2$ this is obvious. Assume that it has been proved for all distributive lattices where $k \leq n$ and let L be a lattice with $n + 1$ elements. We note that a distributive lattice is modular. Therefore it is sufficient that the sub-modular inequality is satisfied on sub-lattices of the form x^+. We know that he lattice is sub-lattice of a product of two totally ordered lattices. Such a product lattice is planer in the sense that it has a Hasse diagram without intersection lines. The Hasse diagram consists of small squares each representing a sub-lattice of the form x^+.

Now consider a polymatroid function h on the lattice. Assume that h is an extreme point in the set of all polymatroid functions. For each point in the lattice the value is constrained by a number of inequalities and since we have assumed that the function is an extreme point at least one of the inequalities holds with equality. We start

at the double irreducible element b. It is contained by two monotone inequalities and one submodular inequality.

$$\begin{aligned} h(a) &\leq h(b) \\ h(b) &\leq h(d) \\ h(a)+h(d) &\leq h(b)+h(c). \end{aligned} \tag{2}$$

The submodular inequality implies that $h(b) \geq h(a) + (h(d) - h(c))$ which a stronger condition than the first monotone condition. Therefore the conditions on $y_1$are

$$h(a)+h(d)-h(c) \leq h(b) \leq h(d). \tag{3}$$

Observe that $h(a) + h(d) - h(c) \leq h(d)$. Since both the lower bound on $h(b)$ and the upper bound on $h(b)$ are linear any extreme polymatroid is also an extreme polymatroid when it is restricted to the lattice where the element b has been removed. According to the induction hypothesis such an extreme polymatroid function only takes the values 0 and 1. Therefore if the polymatroid function on the original lattice is extreme one of the inequalities in (3) must hold with equality and therefore $h(b) = 0$ or $h(b) = 1$. is entropic.

Since an extreme polymatroid function only takes the values 0 and 1 the lattice generated by the polymatroid function has only two elements and this is obviously entropic.

If the lattice is three dimensional one have to modify the above procedure. A three dimensional distributive lattice may not have any double irreducible elements. If a single element is deleted from the lattice it is no longer modular, but modularity is needed if we should use Lemma 12. Instead one consider sequences $(a_1, a_2 \dots a_n), (b_1, b_2 \dots b_n), (c_1, c_2, \dots c_n)$, and $(d_1, d_2 \dots d_n)$ with the conditions

$$\begin{gathered} h(a_j)+h(d_j)-h(c_j) \leq h(b_j) \leq h(d_j) \\ h(d_{j+1})-h(d_j) \leq h(b_{j+1}) - h(b_j) \leq h(b_{j+1})-h(b_j). \end{gathered}$$

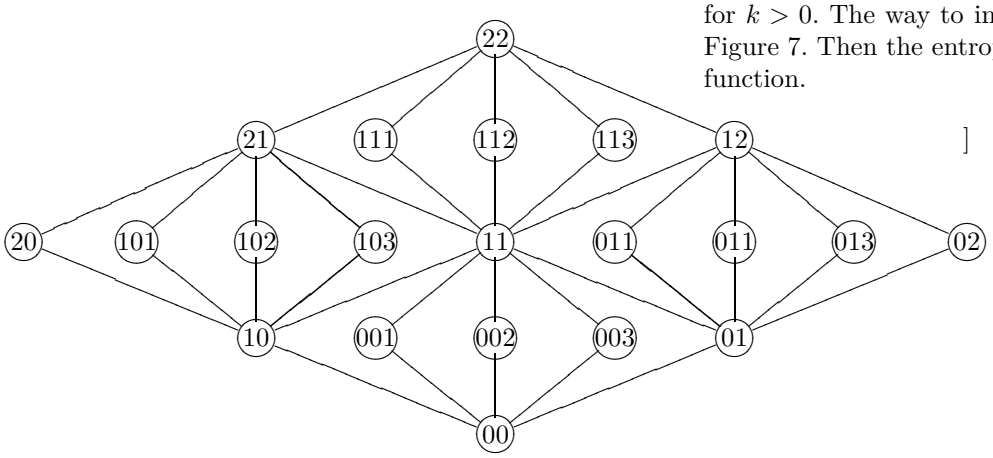

Figure 7. A planar modular lattice with indexing of the elements.

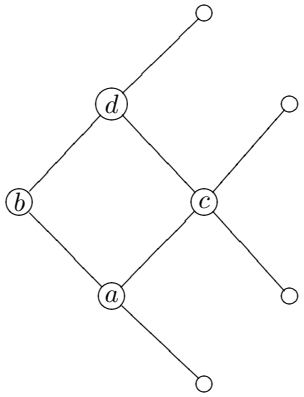

Figure 8. The upper right corner of the lattice.

One can then prove that any extreme polymatroid function only takes the values 0,1/2, and 1 by a more complicated induction argument. One can then use 7 to include that any extreme polymatroid function

Appendix H
Proof of Theorem 17

We use that it has it was recently proved that a planar modular lattices can be represented as a distributive lattice with a number of double irreducible elements added [17]. The proof has the same structure as for distributive lattices, but the existence of the double irreducible elements implies that there are also extreme polymatroid functions that are proportional to the ranking function. Let $X_1, X_2, \dots X_m, Y_1, Y_2, \dots, Y_n$ denote independent random variables uniformly distributed over $\mathbb{Z}_p$ for some large value of p. Let Z_{ij} denote the random variable

$$\biguplus_{\ell \leq i} X_\ell \uplus \biguplus_{\ell \leq j} Y_\ell.$$

and let Z_{ijk} denote the random variable

$$\biguplus_{\ell \leq i} X_\ell \uplus \biguplus_{\ell \leq j} Y_\ell \uplus (X_{i+1} + k \cdot Y_{j+1})$$

for $k > 0$. The way to index the variables can be seen in Figure 7. Then the entropy is proportional to the ranking function.

]